DOI:
10.1080/08351813.2016.1199088

Document Version
Peer reviewed version

Embedded Reference: Translocating gestures in video-mediated interaction

Paul Luff, Christian Heath, Naomi Yamashita, Hideaki Kuzuoka , Marina Jirotka

Abstract:

Audio-visual technologies can enable informal communication akin to face-to-face interaction. However, they prove less successful when deployed to support work and organisational activities. This is, in part, due to the limited ways such technologies provide access to the materials, objects, documents and the like, that are critical for supporting work activities as they emerge and unfold. What is often neglected in these new technologies is a consideration of how objects are referred to, manipulated and transformed within and through interactions between colleagues. In this paper, we consider an advanced prototype system called t-Room that seeks to provide geographically dispersed participants with rich and varied access to physical and digital documents. This technology has been designed to support flexible collaborative activities with and around objects, as if remote participants and materials in their local environment were co-present within a common space. By undertaking quasi-naturalistic experiments in this prototype environment we reveal that at times participants could unproblematically refer to detailed features of the environments and when there were difficulties resolve them. We notice, however, that at other times participants had great difficulties in assessing the relationships between themselves, their remote colleagues and objects in the environment; the very flexibility of the technology introducing instabilities into the sequential accomplishment of referential activities. By considering examples of this technology in use, we suggest that these limitations may reflect wider issues concerned with our understanding of how interactional activities are embedded within the local environment. Data in this paper are in English.

Introduction

The recent emergence of Skype and related video-mediated communication systems, like Google Hangouts, were foreshadowed in the late 1980s and early 1990s when a number of advanced research laboratories around the world started developing so-called 'media spaces' (Fish, Kraut, Root, & Rice, 1992; Gale, 1994; W. W. Gaver et al., 1994; Harrison, 2009). These systems drew on audio-visual infrastructures that it was hoped would be deployed throughout organizations and would transform everyday work and organization. Media spaces largely remained as prototypes within laboratories with very limited deployment into wider everyday working environments (cf. Harper & Carter, 1994). One of their principal shortcomings was that they provided limited resources for supporting work activities. Underpinning their design, and shared by contemporary video-mediated communication systems, was a focus on a 'face-to-face' model of interaction and on 'informal interaction', with little attention being paid to how support could be given for participants to access other kinds of resources for their work. In this paper we consider developments that have sought to overcome these shortcomings; shortcomings that have proved more of a challenge than envisaged by the early developers of media spaces.

The materials discussed in this paper are drawn from video-recordings of participants, interactions and activities through an advanced prototype system, called t-Room. T-Room, as the name suggests, is a room-sized space which presents real-time, life-sized images of remote participants in the environment of the other, as if the participants are sharing the same space. The system consists of two similar environments, based in different physical locations, and aims to enable the participants in each of those environments, to see and talk to each other as if they were trans-located into each other's presence. Analytically we focus on one particular issue, an issue that has proved highly challenging for those with an interest in the design of

such systems, how do we enable participants to satisfactorily refer to and discuss objects and artefacts located within their respective environments.

Like most prototype systems, t-Room is not at a stage of development where it has been deployed to support everyday work activities. Hence, the materials discussed in this paper are drawn from video-recordings of 'quasi-naturalistic experiments', experiments where participants were asked to engaged in open-ended tasks based on activities undertaken in ordinary organisational settings. These tasks require little intervention from those participating in the experiment and were designed to encourage participants to refer to and discuss objects within the respective milieux.

The technology, therefore, offers a novel way of considering issues that have been of interest to a range of scholars interested in embodied interaction: how features of the environment are referred to, animated and manipulated within an ongoing interaction (C. Goodwin, 2000, 2003; Hindmarsh & Heath, 2000; Koschmann, LeBaron, Goodwin, & Feltovich, 2011; Mondada, 2007; K. Murphy, 2004; Keith Murphy, 2005; J. Streeck, 1996; Jurgen Streeck, Goodwin, & LeBaron, 2011). These studies reveal, how referential activities, deixis, 'pointing' and the like are situated, collaboratively achieved and shaped from moment-by-moment by the participants in the course of their production. T-Room aims to provide a coherent environment for such activities and yet, being distributed, needs to break them apart in some way. In this paper we consider the use of t-Room in an experimental setting and how the resources it provides can support participants when they refer to objects, and often quite fine details of objects within interaction. We then consider how on other occasions, despite its sophisticated capabilities, evoking objects becomes problematic within t-Room, and how participants attempt resolve those difficulties. We note that despite the fidelity in which actions appear, the aim to provide similar resources for all participants and resolve previous difficulties with video-mediated technologies, the configuration of the system can undermine

the sequential accomplishment of embodied conduct within an environment. We conclude with a discussion on how experiments with an exotic technology might suggest how we might develop richer understandings of everyday conduct in the environments in which they occur.

Distributing Referential Activity

By providing audio-visual access to a remote colleague, media spaces developed in the early 1990s seemed to offer a rich environment to undertake work activities. They would allow any member of an organisation to immediately set up a link to any colleague and, if the connection was kept on, the media space would provide a sense of co-presence, akin to a remote 'office share' (W. W. Gaver et al., 1992). Curiously, the resources necessary for supporting collaborative tasks, such as those with and around documents, physical objects and other features of the environment were disregarded (C. C. Heath & Luff, 1993). Although additional capabilities for sharing electronic documents were introduced, these were divorced from the audio-visual connection; it was therefore difficult for participants to tie their own conduct with objects and the visual conduct of colleagues. Even when participants explicitly referred to the location of an object, problems emerged, principally because it was hard for a participant to assess the standpoint of their colleague with respect to the features of the local environment (C. C. Heath & Luff, 1992). In part this was due to the limited access provided through video (cf. Keating, Edwards, & Mirus, 2008; Keating & Mirus, 2003), but more it was difficult to assemble coherent sequences of activities with respect to material features of the local environment: the environment of interaction became fragmented and fractured (Luff et al., 2003).

These problems persist in the contemporary applications that are popular today, such as Skype and Google Hangouts, in video conversation undertaken through mobile phones, and in workplace settings where video has been deployed (Licoppe, 2015; Licoppe & Morel, 2009,

2013, 2014a, 2014b). Participants frequently have to undertake quite complex activities in order to 'show' a colleague an object: the referential activity frequently then disrupting the ongoing accomplishment of other tasks or the interaction itself.

Some researchers have sought to develop technologies to provide greater access to material objects within video-mediated interaction. Computer scientists and engineers have developed systems that aim to provide something more akin to working on a common workspace, such as on a tabletop or a shared screen (Ishii & Kobayashi, 1992; Kuzuoka, Yamashita, Yamazaki, & Yamazaki, 1999; Minatani, Kitahara, Kameda, & Ohta, 2007; Tang & Minneman, 1991). So, in addition to providing common access to objects, the Clearboard and Agora systems also project a remote participants' hands into the local domain (Ishii & Kobayashi, 1992; Kuzuoka et al., 1999). This feature allows participants to refer to details of documents, and to contribute in different ways to the tasks at hand (Luff, Heath, Kuzuoka, Yamazaki, & Yamashita, 2006). In particular, through a combination of views, participants can gain a sense of a trajectory of conduct and co-ordinate their own actions with those of a colleague. However, these capabilities rely on participants maintaining a relatively restrained location at a desk. Also, as the participants have to sit on opposite sides of a desk, only two people can collaborate in the distributed space at any time.

In order to consider more flexible ways of supporting collaboration around objects, researchers in Japan developed a system called t-Room (Hirata et al., 2008). Rather than just the hands of the remote participants being projected into the local domain, life size images of the co-participants are presented. These are displayed, in real-time, on a wall of large monitors (called 'monoliths') surrounding the participants (see Figure 1).

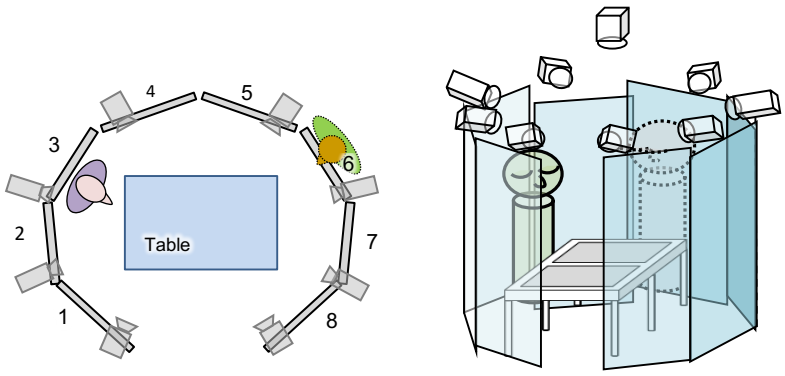


Figure 1: Left: Plan of t-Room showing the position of 8 vertical screens (monoliths). For ease of exposition the monoliths are numbered 1-8: Right: an isometric sketch of the t-Room.

When connected to another t-Room participants in each space have symmetrical resources. Several participants (typically four) can work within this distributed space: the aim being that a remote participant appears as if they are translocated inside the other space. T-Room is also designed to support collaboration with both electronic and physical objects. In the centre of each t-Room there is a tabletop system. When documents are placed on the tabletop they are projected into the other domain. They can also collaborate with electronic images displayed on the walls (the monoliths, see Figure 2).

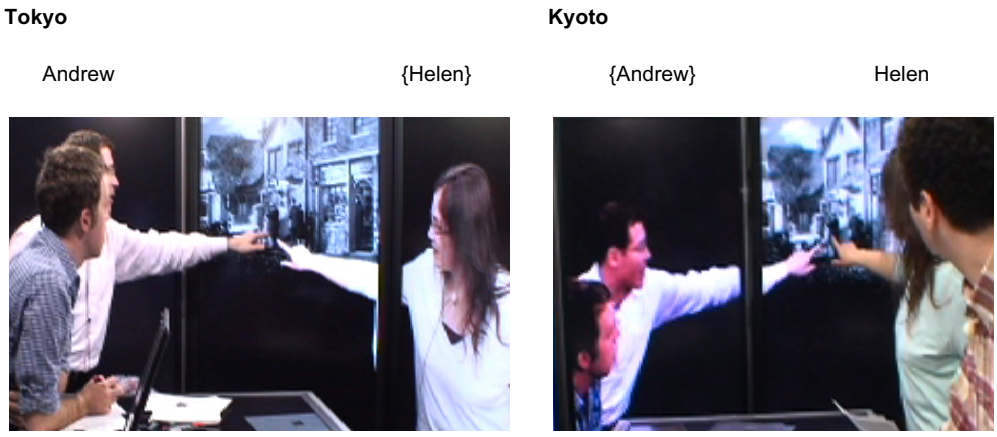


Figure 2: t-Room in two sites: Tokyo (left) and Kyoto (right). All participants are oriented to the same scene of a street displayed on one of the monoliths. Two of the participants, one at each site, are pointing towards this scene. Andrew in Tokyo (on the left) points to a man

in the scene. At the same time Helen in Kyoto points to the same man. In this and subsequent transcripts, images are labeled with the site from which they were taken from (Tokyo or Kyoto) and remote participants (i.e. those in the other site) are identified in{curly brackets}.

The technology is designed to provide a space where people in different places work as if they were together, replicating as far as possible their location and positioning. So, for example, in scientific domains, the system could support the collaborative analysis of complex images, scans, x-rays and the like (Jirotka et al., 2005) or in the social sciences where participants analyse video materials (Tutt & Hindmarsh, 2011) or in design settings where remote teams work on plans, sketches and models (Büscher, Mogensen, Shapiro, & Wagner, 1999). In such settings, participants need to identify objects, and features of objects and also, when interpreting the detail of a scene, object or image, animate their conduct in different ways − through their talk, bodily conduct and gaze direction. As they do this they need to assess the ongoing orientation of their colleagues to what they are animating and showing. An environment such as t-Room presents an intriguing domain to investigate technologically mediated interaction.

As t-Room is a large prototype technology requiring a sophisticated telecommunications infrastructure it cannot be assessed in a naturalistic setting. However, as it requires little training and can be used by individuals to undertake tasks that have much in common with everyday work activities it is possible to develop so-called 'quasi-naturalistic' experiments. In such experiments participants are given open-ended tasks that can take between half an hour to a few hours. Little guidance is given on how to accomplish the task and there is only intervention if some failure of the technology prevents the participants from accomplishing the task. Recently, we have undertaken a series of experiments with different configurations of the t-Room systems that have explored how participants engage in focussed activities around a

single object (Luff, Yamashita, Kuzuoka, & Heath, 2011), how temporal and spatial transformations shape such conduct (Luff et al., 2013) and the different formations participants adopt to undertake collaborative activities (Luff, Yamashita, Kuzuoka, & Heath, 2015).

In this paper we will consider one set of experiments undertaken with t-Room, which drew on recent workplace studies of designers (Luff, Heath, & Pitsch, 2009) to consider how participants might collaborate with a range of resources available in many locations, both physical and electronic. The tasks were developed reflect complex activities required within intensive design meetings such as searching materials, exchanging information about those materials, sharing those materials with others, annotating them and presenting and discussing plans and solutions with colleagues. The task involved the selection and arrangement of materials for a museum exhibition and typically took 90 minutes. The participants were free to organise the work as they saw fit and also to work anywhere they liked within the t-Rooms. Four experiments were undertaken each involving four English–speaking participants who were recruited for the study, two people being in each of the two t-Room spaces. To facilitate the experiments the t-Rooms were located in different places in a research laboratory in Kyoto, Japan. For ease of exposition we will identify the two different t-Rooms as 'Tokyo' and 'Kyoto'. In the experiments we collected materials from 6 cameras (3 from each t-Room). These provided access to most activities undertaken with the t-Rooms. Participants gave written consent to use these materials for research purposes and in subsequent publications.

In the sessions, the participants needed to manage a large number of paper and electronic documents within the space, these could be located on the desktop or on the screens around them. Hence, the participants needed to identify objects, distinguish these from similar ones, locate features within them and describe details of them.

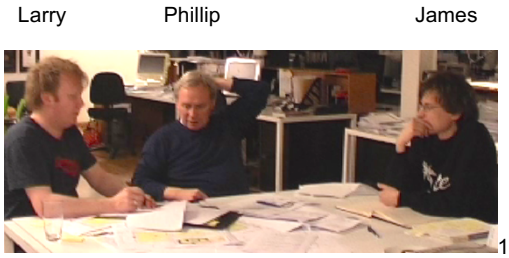
Our analysis is concerned is with the emergent and sequential character of practical action in and through which participants collaboratively accomplish particular activities (C. C. Heath, Hindmarsh, & Luff, 2010). More specifically it draws on materials where the participants engage in referential activities where they, for example, search for appropriate images, make these visible to colleagues, refer to features of documents and recognise these, arrange documents around the space, and plan, discuss and design related textual materials. Although the fragments presented here are short moments from a quasi-naturalistic experiment, they are exemplars of activities that occurred throughout the duration of the tasks. The materials also suggest ways in which participants sought to resolve the problems that occurred, but they also reveal critical ways in which the mediating technology transforms visual conduct.

Embodied referential activity

To give a sense of the kind of activity we will focus on in this paper, we will briefly consider a short fragment of data taken from a study of everyday design practice Three designers of a new museum space are sitting around the corner of a desk and are discussing a particular area of the exhibition for 'classic objects'. Larry (on the left) starts describing this area when Philip (in the centre) asks 'this is not intro? we've done that'.

Fragment 1

Larry: the first area is (0.2) classic objects right? ➔



1.1

(0.3)

Philip: this is ➔

1.2

not intro:? ➔

1.3

we've done that

(0.5) ➔

1.4

Larry: we are here… ➔

1.5

As Larry is starting to describe 'the first area' (1.1), Phillip turns away from Larry, looks down to the desk and moves his hand down to the large document in front of him where there are several sketches of different spaces in the planned exhibition. Following Larry's request for a confirmation ('right?'), Philip looking down at the page (1.2) says 'this is not intro:?' (i.e. not where they are proposing to put the introduction the exhibition). Larry turns first to Philip (1.3) and down to the page, where Philip's hand rests. Phillip then, whilst saying 'we've done that', relaxes his hand and bounces it up and down four times across a small area of the sketch

(1.4). In response, Larry reaches over and points to an area to the right (from his and Philip's viewpoint) of where Phillip's hand has come to rest (1.5), clarifying the location he is referring to with 'we are here'.

Phillip and Larry's collaborative identification of a feature in the document is finely co-ordinated. Philip's turn towards the sketch and his talk secures a re-alignment from Larry to the area in question. Moreover, as is frequently the case, it is not sufficient just to identify the object or feature, but it needs to be described, elaborated or animated in some way. Philip monitors Larry's re-alignment so he can animate the feature on the page which in turn produces a clarification from Larry. Larry, In turn with his index finger outstretched touches a particular area three times as he says 'we are here', slightly adjusting where he is pointing as Philip moves his hand away. Their referential activities are accomplished in concert through sequences of talk, bodily conduct and material actions (C. Goodwin, 2003; Hindmarsh & Heath, 2000). Although apparently simple activities, these are the kinds of conduct that have proven difficult to support through media spaces: in video-mediated communication not only are the domains of the participants and the ecologies around them remote from another, but any gestures made through such technology lose much of their performative impact (C. C. Heath & Luff, 1992). Put crudely, with video-mediated systems it is impossible to 'reach into' a remote domain and shape your actions from moment-to-moment in the light of the conduct of a co-participant (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998).

Convergent Accomplishment of Distributed Embedded Reference

In the course of their activities in the task in t-Room the participants frequently referred to objects on the screens or on the tabletop, not only through their talk but also through their visual conduct, frequently pointing to the monoliths and gesturing over them. So, for example, in the following fragment the participants are just clarifying how many paintings need to be

displayed in each room. Tom in the Kyoto space, is listing the numbers for each room and says there needs to be three in the 'last room'. Gary, also in Kyoto, asks whether he has counted them right.

Fragment 2

Tom:	…the last room is three (paintings).
	(0.7)
Gary:	I was wonder where- where you see three?>Because I see (.) is that not a pain:ting?
	(0.4)
Carl:	no ┌that's the information panel
Tom:	└no that's (for) information
	(.)
Gary:	Oh, Okay Dah Okay

Gary and Tom are positioned at different corners of the desk; Gary looking at a plan on a paper sheet in front of him. Tom, on the other hand, is looking at a similar document that is actually being held by Carl who is in Tokyo. Wendy, the other participant in Tokyo Wendy is walking around the desk to the other side of the room. The document Carl is holding is projected through the desktop and hence is visible to Tom. As Tom says 'last' he points with his right hand above the area he is referring to (2.1K).

Kyoto View

Gary {Wendy} Tom {Carl}

2.1K

↑

Tom: …the last room is three (paintings).

Carl and Tom are looking at the same document through the system and they are in a similar location in their respective sites. For both Carl and Tom in t-Room the image of the other appears behind them on a monolith and so in their current orientation they are barely (or not visible) to each other.

Tokyo View

{Gary| Wendy Carl {Tom}

2.1T

↑

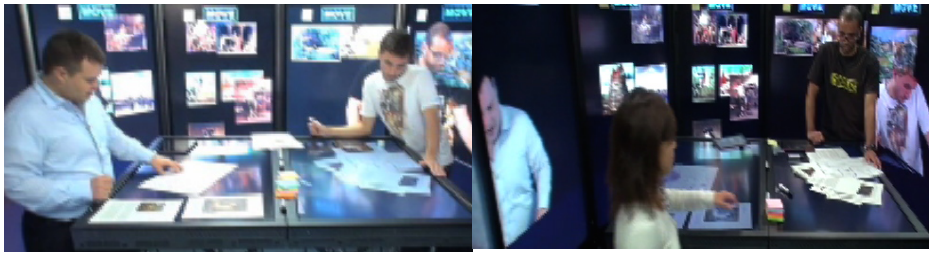
Tom: …the last room is three (paintings).

As Gary asks 'where you see three?' he starts to turn the plan on the tabletop. This turn, accompanied by a bodily movement towards the page, orients the page toward his co-present colleague, Tom (2.2). Because of their positions this conduct and page is also oriented towards Carl. So, as Gary continues with 'Because I see (.) is that not a painting?' Carl looks up towards him (2.3). Gary then turns the page a little more, pointing to a particular location on the bottom right of the page with the first finger of his left hand. As the technology allows the remote participants to see gestures over documents, both Tom and Carl can see the location Gary is referring to. Both Carl and Tom then reorient towards the page and then towards Gary (2.4); Carl also moves 'closer' to Gary (2.4T).

Kyoto **Tokyo**

Gary Tom {Carl} {Gary} Wendy Carl {Tom}

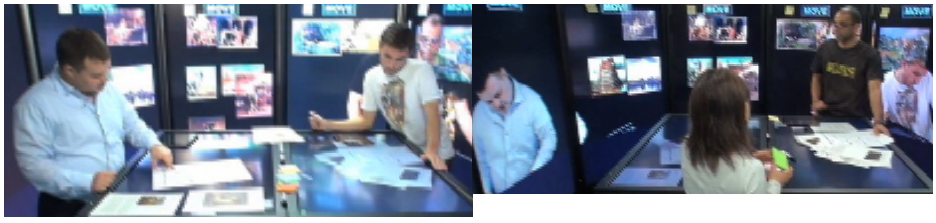
Gary: Because ➔



2.2

I see ➔

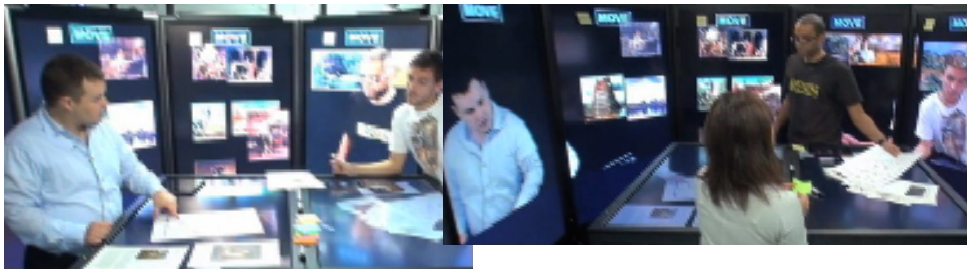
(.)



2.3

is that not a

pain:ting?➔



Gary secures a response from both Tom in the same room and from Carl in Tokyo; both respond, noting that what he is pointing to on the drawing is 'for information' (Tom) or 'the information panel' (Carl) and not a painting. Gary acknowledges this, accompanying his 'Oh, Okay, Dah' with an upward wave of his left hand.

The participants successfully identify that the feature concerned is not a painting but an information panel. Gary manages to locate the item in question on the document, a plan, in front of him, both for a co-present colleague and for one who is remote. This is to a very specific feature – a small drawn object within a region on that plan – and serves to clarify the earlier statement uttered by Tom which refers to the same feature on a similar plan but is in a different space, in fact, it was in the hands of a remote colleague at the time when he mentioned it. This is accomplished without the participants explicitly describing the document they are referring to or explicitly guiding their colleagues to the feature of concern. The participants accomplish this activity through the technology, by drawing on the displayed orientation of their colleagues and the displayed visual conduct of their colleagues. When Gary points, Tom and Carl can draw on his embodied conduct, not only the appearance of his hand but its movement with respect to the document and the reorientation of his body to make sense of his utterance. The clarification is a joint accomplishment by three of the parties in the t-Room that emerges through an interweaving of talk, visual conduct and a material artefact. It relies on one's actions being seen in relation to a prior action by a colleague, and for this to be drawn upon in the next activity. Through this sequential accomplishment of activities, objects, whether remote or in the same space, are embedded within the interaction.

The accomplishment of this referential activity through t-Room would thus seem, in many ways, to resonate with how such activities are accomplished in naturalistic settings, such as in

fragment 1. The technology therefore, seems to facilitate the identification of a feature of a document, the monitoring of a colleagues engagement in that activity and those colleagues producing 'appropriate' responses. Unlike the difficulties found in referential activities in media spaces and other kinds of video-mediated activities, the participants identify and make sense of a detail on a material object within the environment. The object is embedded within the interaction, and its character is critical to making sense of Gary's query. The similarity in both how Carl and Tom respond in the distributed spaces suggests that such conduct can be accomplished through a mediated technology. The participants can not only assess a remote colleague's activities in the light of their own, but can also assume that the ways in which it is being produced remotely is similar to how it is being seen locally.

Managing Incongruent Locations

As well as the documents on the desks being resources for collaborative activities objects displayed on the large screens that surround the t-Rooms were frequently utilized to support the participants perform their tasks. As we join the following example, the four participants are all engaged in different activities, each trying to find particular paintings for the exhibition. The participants in Tokyo (Pete and Mike) are standing on the left hand side of the desk, whilst the participants in Kyoto (Rich and Guy) are on the right. As we join the fragment, Rich, in Kyoto, has been looking across the space for a while at the screen opposite. He then asks what the painting is, that is displayed on that screen.

Fragment 3 (simplified)

➔	R:	whats this one err (1.0) over (0.6) over there?
		(0.5)
	P:	that one?
		(.)

R: Yep

(0.3)

P: the Draughtsman

(.)

R: ah okay

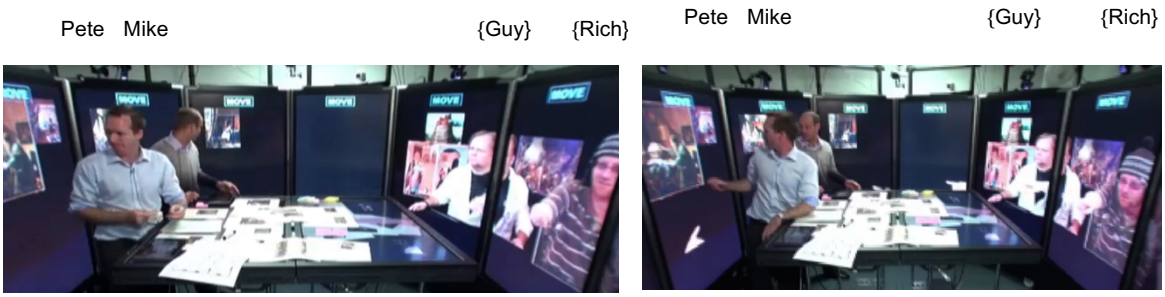
As Rich asks his question he produces a gesture across the desk, pointing with his right hand towards the picture in question (3.1K).

Fragment 3.1 (Kyoto)

Rich's utterance is produced with an 'err', a repetition (over over') and includes two long pauses. Also, his visual conduct also seems perturbed. Before Rich delivers the gist of his utterance he briefly withdraws his arm before going on to produce the gesture again (3.2K), this time reaching out further towards the screen in question (3.3K). Rich's actions are produced in the light of the conduct of his remote colleagues, who seem to be having some problems identifying where Rich is pointing. As Rich begins his utterance both Pete and Mike in Tokyo look up towards him. Seeing Rich's conduct on the screens in front of them, they turn around, each looking towards different screens behind them (3.1T). Mike looks to his left, at the screen where a painting of a woman sat at harpsichord is displayed. Meanwhile, Pete

turns to his right, towards the screen that is directly behind him. This displays an image that was actually added by Rich.

Fragment 3.2 - Tokyo



Pete turns around further (3.3T) and then looks across the screens to his right raising his right arm before arriving at an image on that screen. As he points he says 'that one?' (3.4T) which Rich confirms. Only after he has said this does Rich withdraw his left arm to his side (3.5K)

Fragment 3.3 - Tokyo View **Kyoto View**

Pete then answers the original question with the name of the painting ('the Draughtsman') which seems to identify successfully the picture that Rich is asking about. Despite Pete (and Mike) immediately recognizing the question is for them and Rich's sustained gesture towards

the object, there are difficulties identifying which object is being pointed at. In part, this seems be due to how Rich's conduct appears on the screen.

Although Rich is pointing to the third 'monolith', the screen to Pete's left, for Pete, Rich appears to be pointing at the second monolith, the screen behind him (Pete). This is due to a variation of what is known as the 'Mona Lisa' effect (Anstis, Mayhew, & Morley, 1969). Typically associated with gaze, the 'Mona Lisa' effect is where the video image of a remote colleague appears to be directing their gaze towards a viewer, even when they are looking away. In this case, rather than the direction of gaze appearing misplaced, there are differences between how the production of a gesture is produced and how it appears in the remote environment. For Pete, Rich's conduct, his outstretched arm, appears to be directed towards himself, and not to his left.

Fragment 3 - Detail (Tokyo view)

Hence, in response to the request 'what's this one' Pete turns right around and tries to find an appropriate referent for Rich's point. Similarly, Mike sees the gesture being directed towards himself and turns and looks behind himself. Unlike the previous fragment, Rich's actions engender quite different responses from his two remote colleagues. They both look towards different objects and away from the one Rich is pointing at within his own space.

The conduct of his remote colleagues is visible to Rich, and he adopts quite a conventional practice in order to resolve the difficulty (Goodwin 1981. Following perturbations and a pause, he restarts his utterance and his gesture (Charles Goodwin, 1981). Rich withdraws his hand as Pete starts to turn (3D2), and then extends it again, repeating 'over'. This time he animates his gesture, moving his finger from left to right (3D3), as if to guide Pete towards the correct place. Rich sustains his gesture as Pete continues his search to resolve what Rich is referring to. In the end, Pete does manage to identify the object in question, but this is not just in the light of the repetitive and sustained conduct of Rich but also through an unintended feature of the technology (one of the other cameras captures Rich's outstretched hand and displays this on one of the screens behind Pete – just visible on the bottom left of image 3.3T)

As with more conventional kinds of video-mediated communication, participants' conduct is transformed through the technology (W. W. Gaver, 1992; C. C. Heath & Luff, 1992). Moreover, when producing the gesture Rich has few resources for assessing how his pointing gesture is being understood. He can, in some ways, 'repeat' his conduct, but has little to draw on to re-design or re-shape his conduct that is sensitive to the nature of the problem being faced by his remote colleagues.

Unlike other video-mediated technologies, t-Room provides capabilities for supporting collaboration on and over documents, digital and physical, on both the 'walls' and the 'tables'. These capabilities have been carefully designed so that participants' conduct is not divorced from the objects in question. As seen in fragment 2, when participants are close to an object

on the screen and they point to a feature of it, the appearance of their hand and fingers appears over the object and it is possible to distinguish detailed features and aspects of that conduct. However, as t-Room is also designed to be flexible and is quite large, participants can choose where to position themselves and have a variety of ways of referring to objects. When they do position themselves further away from an object that they refer to, then the video technology transforms how referential conduct appears: it does not appear in the remote domain as it does when produced in the local environment. Participants become divorced from the object they are referring to (cf. Hindmarsh et al., 1998; Luff et al., 2003). Moreover, those producing the actions have few resources for assessing how their own conduct appears, how it has been transformed when trans-located in the remote domain. Thus, a fragility emerges regarding the assumptions they can draw about the relationships between their own conduct and the environment in which it is embedded. Hence, when problems emerge they have few resources for assessing the nature of the problem and designing ways to resolve the ensuing difficulties.

Hybrid Asymmetries of Action

As t-Room allows participants to position themselves in any part of the room, they can adopt a variety of configurations for undertaking collaborative activities over documents. In the following instance they position themselves along one side of the desk. When we join them the participants are trying to find one more painting that will go with the Renaissance pictures which they have already selected. Tim in Tokyo and Charles in Kyoto are looking through a collection of paper documents on the desktop, whilst Runako in Kyoto is looking at a description of one painting on a paper sheet she is holding in front of her. Meanwhile, James in Tokyo rests on the desktop oriented towards Runako and the sheet she is holding. Runako announces she may have a candidate picture to include in the exhibition.

Fragment 4 (simplified)

	R:	I have the one from: (0.7) (it says here) fighting the dragon (.)
		(.)
	J:	ooohh (.) that might be any good (.) do you have that (do you)
		(.)
	R:	yeah
➔	T:	put it=
	R:	=(Blue one)
		(.)
➔	T:	put it down on the table um::
		(.)
	R:	okay
		(.)
	T:	(Runako) so we can see it
		(.)
	R:	okay
		(0.5)
	T:	where?
		(.)
	J:	is he on a horse
		(.)
	T:	where is it?

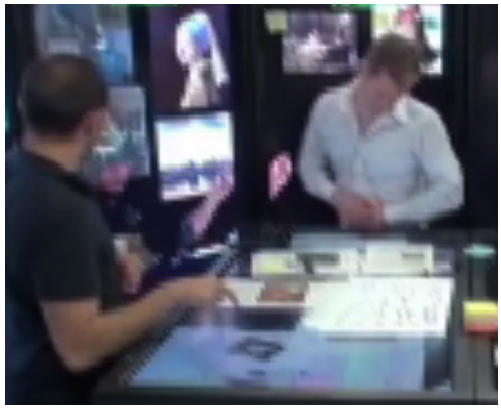
Tim, in Kyoto, who is working by Runako's side turns towards her and suggests that she put the sheet down on the tabletop so they can all see the details ('put it down on the table um:: (Runako) so we can see it') (4.1T). As he says this he holds his right hand over a location in front of him (and to Runako's right) with his index finger pointing down to an area he has just cleared (4.4T).

Fragment 4 (Tokyo View)

Tim {Runako} James

4.1T

Tim {Runako} James



4.4T

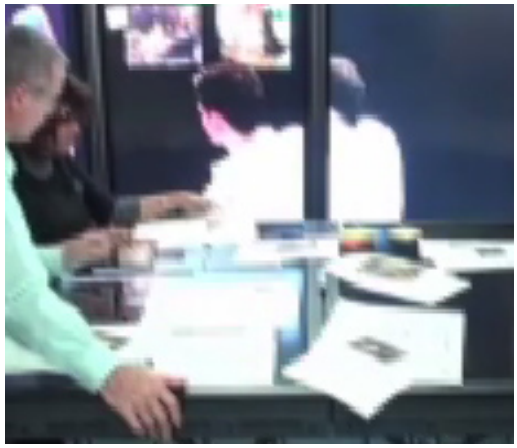
T: put it put it down on the table um:: (Runako) so we can see it

R: (Blue one) okay

4.1K

Charles {James}

Runako



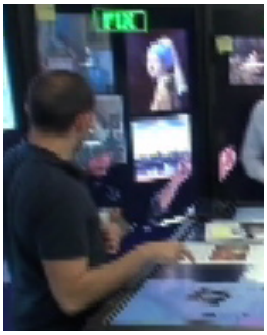
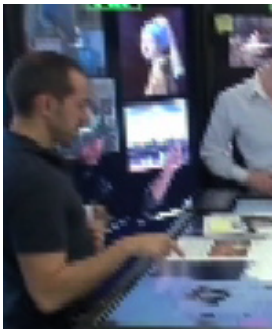
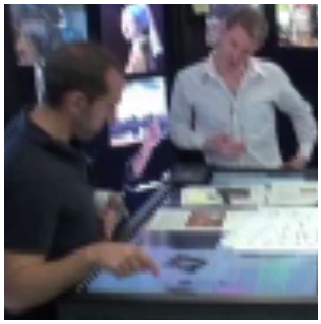
4.3K

Charles {James}

Runako

Runako agrees and does put the sheet down in front of her and slightly to her left (nearer to James than Tim) (4.2K). Tim sees Runako on the monolith to his left put the sheet down (4.5T). He then turns (4.6T) and looks down to the desk in front of him still holding his right hand above a space on the desktop. He then asks 'where is it?' (4.8T).

Fragment 4 (Tokyo View)

Tim {Runako}	Tim {Runako} James	Tim {Runako} James
		
4.5T	4.6T	4.7T
T: so we can see it (.)	R: okay (0.5)	T: Where is it?

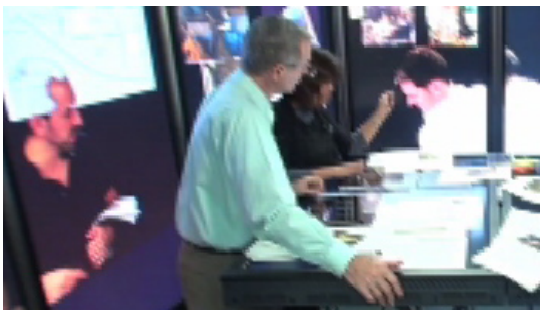
Tim's conduct is carefully designed for Runako. He is oriented to Runako, who appears just to his left, and as he asks her to 'put it down on the table so we can see it' he holds his hand over an 'empty space' on the desk, his finger pointing down (4.3T – 4.7T). It is designed so that Runako, who appears alongside him can see his conduct, but also the nearby location he is referring to. Runako agrees to Tim's suggestion and takes an action to accomplish it, and yet Tim cannot determine the location where she undertook that action. Indeed, Tim looks even further to the right to try and find the document, still maintaining his gesture as he inspects a place at the opposite end of the desk to where Runako has placed it.

In the design of his gesture Tim assumes that Runako can see his conduct and assess its relevance. It indicates a place on a crowded desk where her document would be visible. And yet, in the Tokyo space Tim's actions appear very differently. Tim's image appears behind Runako, and so his actions are not visible to Runako (4.2K). Nevertheless, Tim's colleague in the Kyoto space, James, is visible to Runako's right and he re-orients to the document she is referring to.

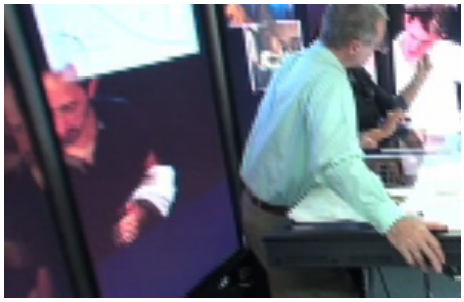
Fragment 4 (Kyoto View)

Charles Runako

{Tim } {James} Charles Runako



{Tim} {James}



4.2K ↑ **4.7K** ↑

T: put it down on the table um:: (.) (Runako) so we can see it (.)

R: Okay Okay (0.5)

Indeed, as Runako places the document down on the desk, James turns to face the desk and looks down (4.6K). So, for Runako, when she puts the document down, she sees James's reorientation in its direction as if in response to her action. Similarly, her co-present colleague Charles also reorients to the document in front of her, moving towards her and it.

In spite of all the participants standing in a 'virtual' row alongside the desk and all being very 'close' to the object in question, they each have quite different perspectives on the action. Most importantly they all have different ways of seeing their colleague's actions in relation to their local ecology. So for example, Tim can see Runako and features of her local environment. He can assess her conduct in relation to his own. However, Runako can see little of Tim, she can hear him but has few resources for tying his conduct (such as his talk) to features of the local environment. Nevertheless, the actions of others, James in Kyoto and Charles alongside her, are visible and also seem sequentially related and responsive to her own. Unlike in fragment 3, when Rich by sustaining, extending and re-productng a gesture, if not resolving a referential problem, at least establishes the foundations for resolving it, here Tim's actions, his production and reproduction of a gesture and his accompanying reorientation and talk, does not contribute to a concerted, collaborative activity. Tim can

undertake remedial actions, but the technology undermines the means he has for making sense of his colleagues' conduct with respect to these.

So, despite all participants appearing to have very similar resources available to them in this video-mediated environment, the technology actually produces different ecologies of action for each participant. Participants have visible access to a remote object, quite fine details of that object and, more importantly, to remote colleagues and how they are orienting to that object from moment-to-moment. They have resources for monitoring the emerging conduct of a remote colleague and for producing their own conduct in relation to their co-participants and the objects around them. That is, they seem to have similar resources to participants in co-present settings to establish coherent sequential activities with objects within an environment. And yet, a small change in position or orientation of a participant can mean that one of these resources can be, even momentarily, unavailable and then the sequential accomplishment of object-related activities can become problematic.

This can mean that seemingly subtle differences where a participant is located or how an action is shaped can have quite marked consequences for how it appears to a remote participant, or whether it is visible at all. Despite the very rich audio-visual environment in which conduct is presented, the sequential accomplishment of embodied action through the technology can rest on quite fragile foundations.

Discussion

> "The individual gestures with the immediate environment, not only with his body, and so we must introduce this environment in some systematic way … while the substratum of a gesture derives from the maker's body, the form of the gesture can be intimately determined by the microecological orbit in which the speaker finds

himself. To describe the gesture, let alone uncover its meaning, we … have to introduce the human and material setting in which the gesture is made. …"

(Goffman, 1964): 164

Video-mediated technologies like media spaces offer rich support for interpersonal communication, as do commercial systems, such as Skype and Google Hangouts. By focussing on facilitating 'face-to-face' interaction facial conduct and some bodily gestures are made accessible and these support many communicative activities, like turn-taking, some iconic gestures and special kinds of 'communicative' gestures (e.g. Keating et al., 2008), These technologies also offer some facilities for remote participants to have greater access to the environment of the other, particularly the objects and artefacts that surround them. However, this access is limited, and usually constrained by having a single camera in a fixed position. Even when a camera is mobile it can hard to establish a relationship between the participants, the viewpoint and the objects in question (Licoppe, 2015; Licoppe & Morel, 2009, 2013, 2014a, 2014b). In order to support richer forms of collaborative work, even just being able to refer easily to a particular feature of an object, requires more sophisticated solutions that offer greater access to a remote environment and to the conduct of the other.

Technologies that aim to offer rich forms of real-time distributed co-operative work also seem to offer the potential of addressing problems faced by contemporary dispersed organisations, whether these are to enhance collaboration between participants in different continents or to support participants to undertake tasks and activities remotely. Sophisticated techniques have been developed to provide an accurate sense of the direction of the gaze of a remote participant and to preserve the shifts in orientation that occur when participants are co-present so that these are consistent in local and remote spaces (O'Hara, Kjeldskov, & Paay, 2011). However, even in these advanced systems, the material setting, the environment of objects and artefacts, in which interactions are accomplished tend to be neglected. The focus

of attention in technological development tends to be on producing higher fidelity, face-to-face interactions. Certain gestures can be re-presented in fine detail, but like the social scientists Goffman criticised, developers of such systems seem to neglect that the 'form' of the gesture is 'determined by the microecological orbit in which the speaker finds himself'. They, too, neglect the situation in which visual conduct is accomplished (Goffman, 1964).

Providing mediated access to the material environment, however, is not a mere matter of providing larger and greater numbers of screens and more cameras with which to capture and display features of the environment. In the example considered here, t-Room, the system provided access to a rich array of heterogeneous objects, including large images and documents such as plans through to small pictures, sketches and reports. Participants could place these locations that were appropriate to the tasks they wish to accomplish, whether this was on a surface like a wall or a desk. They could also arrange themselves in a variety of ways in the space in order to accomplish collaborative tasks (Luff et al. 2015), so they could discuss the detail of a document, compare one object with another or make a proposal to a number of colleagues. What seemed critical when accomplishing these collaborative activities was not just that the participants had visual access to an object, but also they had access to their colleague's conduct in relation to it; how they were orienting to that artefact and how a colleague's activities were produced in the light of one's own. In other words, the technology provides the resources to support the emergent, concerted and sequential accomplishment of collaborative activities with objects. When undertaking an activity such as when one participant refers to a detail of an object though a gesture, they can draw on the orientation and re-orientation of their co-participant towards that object and can shape their own conduct in the light of the actions of the co-participant. An activity such as when one person points to an object on a screen can then be 'geared' with respect to the shifting participation of another, so a feature is identified just when the other orients to it. The technology allows

participations to project trajectories of conduct so these can then be developed upon in subsequent activities (cf. Luff et al. 2011, 2013). The object is embedded within a collaborative and sequential course of activities. By presenting rich 'embodiments' of the participants, not just the 'head and shoulders', but often life-sized presentations of the other in 'real-time', a video-mediated technology can seem to provide resources for identifying, referring and discussing objects in interaction. For both the participant initiating the referential activity and the 'recipient(s)' features of the material environment seemed to be embedded unproblematically within the unfolding interaction in ways that resonate with how objects are referred to in everyday interaction focused around artefacts (Hindmarsh & Heath, 2000; Nevile, Haddington, Heinemann, & Rauniomaa, 2014; Jurgen Streeck et al., 2011).

And yet, on other occasions what appear to be very similar referential activities could seem fragmented and disjointed. For example, when a participant undertook an activity like pointing to a distant object it seemed evident that one or more colleagues were having difficulties making sense of their conduct. The ways in the participants addressed such difficulties revealing not only ways in which participants sought to remedy the problems but provide analysts with access to what these difficulties were. Participants would tend to 'repair' the conduct in question by maintaining, extending or even restarting the problematic action, and if this failed to secure alignment, produce a more explicit kind of referring activity, even directing another to the feature just through talk. Through this conduct they revealed the differing perspectives the participants have within the video-mediated environment. In common with other kinds of video-mediated communication asymmetries are introduced, how an action appears in the local environment can be different when translocated into the remote one. These transformations are consequential for how interactions are accomplished (e.g. Keating et al., 2008) particularly when features of the material environments are invoked (C. C. Heath & Luff, 1992). A local participant can have limited resources for assessing how

their visual conduct appears to other participants. What appears distinctive about t-Room is that these asymmetries are themselves unstable. So, when participants are positioned near to some common location, say when they are all oriented towards an object on one of the large screens, the orientations and movements of the local and the remote participants (represented through life size images) can be drawn on to assess the standpoint of another within the local environment. However, when a participant turns towards another object, say on the desk, even despite the remote colleague (and their life-sized image) similarly turning in response, the interactional environment can shift dramatically. A participant can no longer rely on a remote colleague having similar access to their own conduct and the constellation of resources around them. Despite trying to develop a technology that aims to reproduce, as close as possible, co-present referential conduct, the designers have developed an unstable environment for interaction with and around objects.

The designers of t-Room carefully considered how to develop a system that could provide a very rich and flexible environment for undertaking collaborative activities. Subtle changes in a participant's body movement, gaze direction and aspects of their gestures and bodily conduct are visible to colleagues who may be hundreds of kilometres away. It also offers almost unprecedented access to a range of kinds of object, both physical and digital, which participants can work on in a variety of locations and orientations. The participants can also shift between different kinds of activity, from working on individual tasks, to briefly clarifying an aspect or feature of an artefact to discussing objects with several colleagues, both local and remote, at the same time. In contrast to other 'high fidelity blended environments' (e.g. O'Hara et al., 2011), this is facilitated by t-Room allowing participants to move around the space so that they can establish different configurations of how they position themselves with respect to another and the objects that are available in the local and remotes spaces (Luff et al

2015). However, this flexibility means that the relationships between a participant, an object and a co-participant can be uncertain.

With current video systems, asymmetries are unavoidable between how conduct is seen in the local environment and how it is seen from a remote standpoint. The relationships between a participant, a remote colleague and any objects in the environment are transformed through the mediating technology. If these asymmetries are stable then participants can produce an action, monitor the ongoing activities of a co-participant and then can shape and, if necessary take remedial action and reshape that conduct in the light of their co-participant's participation. In the case of t-Room its very flexibility led to different asymmetries being invoked when the spatial relationships between the participant, the co-participant and the object changed in often very subtle ways. In such cases, participants had little way of telling that their conduct was now not visible in the way it was produced. The interactional asymmetries were themselves unstable, changing after just a small alteration of position or in orientation. It then proved difficult for participants, from moment-to-moment to assess the standpoint of another with respect to an object in the local environment. At these times participants could not make use of the local environment to make sense of the conduct of their colleagues, and had no systematic way to identify a probkem and to shape any remedial action.

In recent years, designers of collaborative technologies have developed some very sophisticated technologies that offer unprecedented support for remote interpersonal communication and that allow access to rich kinds of resources when participants are geographically dispersed. Developments in video-mediated technologies do offer ways of identifying, referring to and discussing remote objects, but these tend to rely on either the participants or the objects remaining in a relatively stable location, or both. This may not be that surprising. In the social sciences, there has been a longstanding interest in considering

embodied actions, particularly gestures, with respect to communication and conversation (Argyle & Cook, 1976; Kendon, 1990, 2004). Whilst directing attention towards the body it may also divert it away from the objects that are critical resources for social action. Recent detailed and insightful studies by social scientists have started to reveal how objects are used, manipulated and otherwise evoked in social interaction. These pay particular attention to how objects feature as resources in social interaction, how they are critical in making sense of ongoing talk, how they serve to co-ordinate collaborative actions by colleagues and how they can be invoked as resources to achieve a range of social activities, whether these to assist in the instruction of others (Mondada, 2014), to delegate actions to colleagues (Weilenmann & Lymer, 2014) or even serve to define and discover the precise nature of an object (Koschmann & Zemel, 2014). But even in these studies there is a tendency to focus on occasions where the relationships with objects are fairly constrained; either by focusing on a single object, objects in a circumscribed domain or where the standpoints between participants and the objects are fairly stable. There has been less attention about how participants draw on assemblies of artefacts, juxtapose objects in different domains or where those objects are in less fixed locations, in other words, how the objects are embedded within an environment, within the 'microecological orbit' of other objects and other co-participants. This is understandable, as not only are such activities less amenable to analysis but they also set methodological challenges for analysts for gathering data, analysing these and subsequently presenting the analysis. The technology considered here may seem rather exotic, its designers aiming to provide a means of mediating complex, collaborative activities with different kinds of object on different surfaces between participants who would be a great distance away from each other. Experiments with this technology do, however, reveal are that we rely from moment-to-moment on the occasioned features of the environment to make sense of another's conduct,

and yet the practices through we embed everyday social actions so intimately within the local ecology are still matters that deserve our close scrutiny.

Acknowledgements

We would like to thank members of the NTT t-Room team and the WIT Research Centre for their help and support.

References

Anstis, S. M., Mayhew, J.W., & Morley, T. (1969). The Perception of Where a Face or Television 'Portrait' Is Looking. *American Journal of Psychology, 82*(4), 474-489.

Argyle, M. , & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge Cambridge University Press.

Büscher, M., Mogensen, P., Shapiro, D., & Wagner, I. (1999, 12 - 16 September). *The Manufaktur: Supporting work practice in (landscape) architecture.* Paper presented at the ECSCW '99, Copenhagen.

Fish, R. S. , Kraut, R. E. , Root, R. W. , & Rice, R. E. (1992, 3rd-7th May). *Evaluating Video as a Technology for Informal Communication.* Paper presented at the CHI '92, Monterey, CA.

Gale, S. (1994). Desktop Video Conferencing: Technical Advances and Evaluation Issues. In S. A. R. Scrivener (Ed.), *Computer-Supported Cooperative Work: The multimedia and networking paradigm* (pp. 81-104). Aldershot: Avebury Technical.

Gaver, W. W. (1992, Oct 31 - Nov 4). *The Affordances of Media Spaces for Collaboration.* Paper presented at the CSCW '92, Toronto, Canada.

Gaver, W. W. , Moran, T. , Maclean, A. , Lovstrand, L. , Dourish, P. , Carter, K. A. , & W., Buxton. (1994). Working Together in Media Space: CSCW Research at EuroPARC.

In S. A. R. Scrivener (Ed.), *Computer-Supported Cooperative Work: The multimedia and networking paradigm* (pp. 185-205). Aldershot: Avebury Technical.

Gaver, W. W. , Moran, T., Maclean, A. , Lovstrand, L. , Dourish, P. , Carter, K. A. , & Buxton, W. (1992, 3 - 7 May). *Realizing a video environment: EuroPARC's RAVE system.* Paper presented at the CHI 92, Monterey, CA.

Goffman, E. (1964). The Neglected Situation. *American Anthropologist, 6*(2), 133-136.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*, 1489-1522.

Goodwin, C. (2003). Pointing as a Situated Practice. In S. Kita (Ed.), *Pointing: Where Language, Culture and Cognition Meet* (pp. 217-241). Mahwah, NJ: Lawrence Erlbaum.

Goodwin, Charles. (1981). *Conversational Organisation: Interaction between Speakers and Hearers*. London: Academic Press.

Harper, R., & Carter, K. (1994). Keeping People Apart. *CSCW Journal, 2*(3), 199-207.

Harrison, S. (Ed.). (2009). *Media Space 20 + Years of Mediated Life* London: Springer-Verlag.

Heath, C. C. , Hindmarsh, J., & Luff, P. (2010). *Video in Qualitative Research: analyzing social interaction in everyday life*. London: Sage.

Heath, C. C. , & Luff, P. (1992). Media Space and Communicative Asymmetries: Preliminary Observations of Video Mediated Interaction. *Human-Computer Interaction, 7*, 315-346.

Heath, C. C., & Luff, P. (1993, June). *Media Space and Communication Asymmetries: Privacy and Control within an Organisation Environment.* Paper presented at the Espaces publics: esthetiques de la democratie, Cerisy-la-Salle, France.

Hindmarsh, J., Fraser, M., Heath, C. C., Benford, S., & Greenhalgh, C. (1998). *Fragmented Interaction: Establishing mutual orientation in virtual environments.* Paper presented at the CSCW'98, Seattle, WA.

Hindmarsh, J., & Heath, C. (2000). Embodied Reference: A Study of Deixis in Workplace Interaction. *Journal of Pragmatics, 32*(12), 1855-1878.

Hirata, Keiji , Harada, Yasunori, Takada, Toshihiro , Aoyagi, Shigemi, Shirai, Yoshinari , Yamashita, Naomi , . . . Nakazawa, Kenji. (2008). *t-Room: Next Generation Video Communication System* Paper presented at the World Telecommunications Congress at IEEE Globecom (WTC'08).

Hsieh, Gary, Wood, Kenneth, & Sellen, Abigail. (2006). *Peripheral display of digital handwritten notes*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montreal, Que;bec, Canada.

Ishii, H., & Kobayashi, M. (1992). *ClearBoard: a seamless medium for shared drawing and conversation with eye contact.* Paper presented at the CHI'92.

Jirotka, M., Procter, R., Hartswood, M., R., Slack., Simpson, A., Coopman, C., . . . Voss, A. (2005). Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study. *Computer Supported Cooperative Work, 14*(369-398).

Keating, E., Edwards, T., & Mirus, G. (2008). Cybersign and new proximities: Impacts of new communication technologies on space and language. *Journal of Pragmatics, 40*, 1067-1081.

Keating, E., & Mirus, G. (2003). American Sign Language in virtual space: Interactions between deaf users of computer-mediated video communication and the impact of technology on language practices. *Language in Society, 32*, 693-714.

Kendon, A. (1990). *Conducting interaction: Studies in the Behaviour of Social Interaction*. Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Koschmann, T., LeBaron, C., Goodwin, C., & Feltovich, P. (2011). "Can you see the cystic artery yet?" A simple matter of trust. *Journal of Pragmatics, 43*, 521-541.

Koschmann, T., & Zemel, A. (2014). Instructed Objects. In M. Nevile, P. Haddington, T. Heinemann & M. Rauniomaa (Eds.), *Interacting with Objects: Language, materiality and social activity* (pp. 357 - 380). Amsterdam: John Benjamins Publishing Company.

Kuzuoka, H., Yamashita, J., Yamazaki, K., & Yamazaki, A., A. (1999). *Agora: A Remote Collaboration System that Enables Mutual Monitoring.* Paper presented at the CHI'99 Extended Abstracts, Philadelphia PA.

Licoppe, C. (2015). Video communication and 'camera action'; The output of video shots wide in courtrooms with remote Defendants. *Journal of Pragmatics, 76*, 117-134.

Licoppe, C., & Morel, J. (2009). *The collaborative work of producing meaningful shots in mobile video telephony.* Paper presented at the Mobile HCI'09.

Licoppe, C., & Morel, J. (2013). Appearings in Video Communications: Interactionally generated encounters and the accomplishment of mutual proximity in mobile phone conversations. In P. Haddington, L. Mondada & M. Nevile (Eds.), *Interaction and Mobility: Language and the Body in Motion* (pp. 277-299). Berlin: De Gruyter.

Licoppe, C., & Morel, J. (2014a). *Appearings in Video Communications.* Paper presented at the Skype Connections and the Gaze of Friendship and Family, Cambridge UK.

Licoppe, C., & Morel, J. (2014b). Mundane video directors in interaction: Showing one's environment in Skype and mobile video calls. In M. Broth, E. Laurier & L. Mondada (Eds.), *Studies of Video Practices: Video at Work* (pp. 135-160). Abingdon, Oxon UK: Routledge.

Luff, P., & Heath, C. (2000). The Collaborative Production of Computer Commands in Command and Control. *international Journal of Human-Computer Studies, 52*, 669-699.

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured ecologies: creating environments for collaboration. *Special Issue of the HCI Journal : 'Talking About Things: Mediated Conversations about Objects', 18*(1-2), 51-84.

Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K., & Yamashita, J. (2006). *Handling Documents and Discriminating Objects in Hybrid Spaces* Paper presented at the CHI 2006, Montreal.

Luff, P., Heath, C., & Pitsch, K. (2009). Indefinite precision: the use of artefacts-in-interaction in design work. In C. Jewitt (Ed.), *Routledge Handbook of Multimodal Analysis* (pp. 213-222). London: Routledge.

Luff, P., Jirotka, M., Heath, C., Eden, G., Yamashita, N., & Kuzuoka, H. (2013). Embedding Interaction: the accomplishment of actions in everyday and video-mediated environments. *ACM Transactions on Computer-Human Interaction, 20*(1).

Luff, P., Yamashita, N., Kuzuoka, H., & Heath, C. (2011). *Hands on Hitchcock: Embodied Reference to a Moving Scene,*. Paper presented at the CHI 2011, Vancouver.

Luff, P., Yamashita, N., Kuzuoka, H., & Heath, C. (2015). *Flexible Ecologies And Incongruent Locations.* Paper presented at the CHI 2015,, Seoul, South Korea.

Minatani, S., Kitahara, I, Kameda, Y. , & Ohta, Y. (2007). *Face-to-Face Tabletop Remote Collaboration in Mixed Reality.* Paper presented at the ISMAR'07, IEEE Comp. Soc.

Mondada, L. (2007). Operating together through videoconference: Members' procedures for accomplishing a common space of action. In S. Hester & D. Francis (Eds.), *Orders of Ordinary Action* (pp. 51-67). Aldershot: Ashgate.

Mondada, L. (2014). The shaping of things in the kitchen. In M. Nevile, P. Haddington, T. Heinemann & M. Rauniomaa (Eds.), *Interacting with Objects: Language, materiality and social activity* (pp. 199 - 226). Amsterdam: John Benjamins Publishing Company.

Murphy, K. (2004). Imagination as Joint Activity: The Case of Architectural Interaction. *Mind, Culture, and Activity, 11*(267-278).

Murphy, Keith. (2005). Collaborative Imagining: The Interactive Use of Gestures, Talk, and Graphic Representation in Architectural Practice. *Semiotica, 156*(113-145).

Nevile, M., Haddington, P., Heinemann, T., & Rauniomaa, M. (Eds.). (2014). *Interacting with Objects: Language, materiality and social activity*. Amsterdam: John Benjamins Publishing Company.

O'Hara, Kenton, Kjeldskov, Jesper, & Paay, Jeni. (2011). Blended Interaction Spaces for Distributed Team Collaboration. *ACM Trans. On Computer Human Interaction, 18*(1), 3-3.

Streeck, J. (1996). How to do things with things: objects trouve and symbolization. *Human Studies, 19*, 365-384.

Streeck, Jurgen, Goodwin, Charles, & LeBaron, Curtis (Eds.). (2011). *Embodied Interaction: Language and Body in the Material World*. Cambridge: Cambridge University Press.

Tang, J. C. , & Minneman, S. L. (1991). VideoDraw: A Video Interface for Collaborative Drawing. *ACM Transactions on Information Systems, 9,*(2), 170-184.

Tutt, D., & Hindmarsh, J. . (2011). Reenactments at work: demonstrating conduct in data sessions. *Research on Language & Social Interaction, 44*(3), 211-236.

Weilenmann, A., & Lymer, G. (2014). Incidental and essential objects in interaction: Paper documents in journalistic work. In M. Nevile, P. Haddington, T. Heinemann & M. Rauniomaa (Eds.), *Interacting with Objects: Language, materiality and social activity* (pp. 319 - 337). Amsterdam: John Benjamins Publishing Company.